Put on your producer hat: considerations for FAIR video material

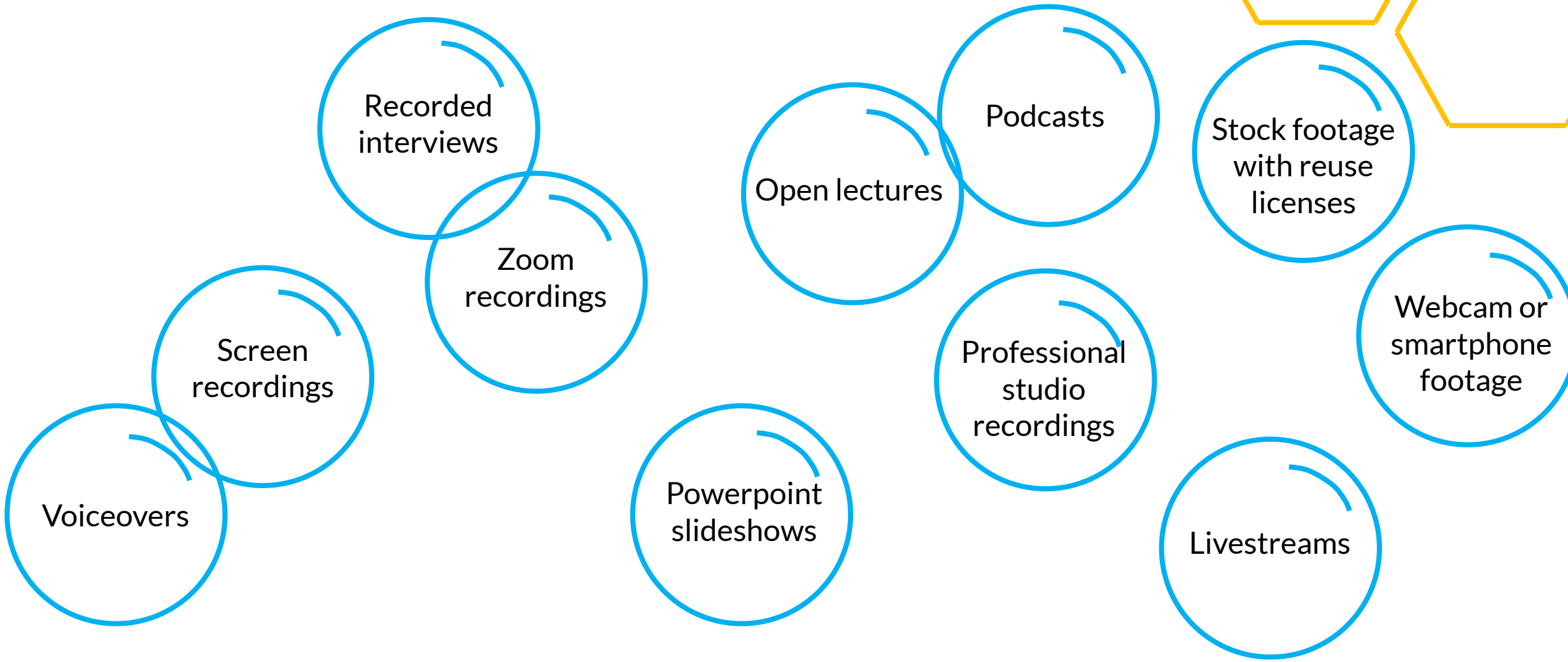
Module 6 – FAIR Training Material by Design
September 18-19, 2024
Kristen Schroeder

SciLifeLab

Learning outcomes for this session

- **Describe** the advantages and challenges of video media in FAIR training
- **Understand** active and responsible AI usage in creating FAIR video materials
- **Find** video materials to use in FAIR training
- **Annotate** video material to make it FAIR
- **Apply** FAIR video materials to your training delivery system

Just like research data, any **publicly funded video training material** should be **FAIR** to enable its **discovery, evaluation & reuse.**



Part 1: Why do we care about having FAIR videos?

Video material…

- assists in asynchronous learning by adding more dimensions to the learning experience
- increases access to knowledge by removing location and time barriers to experts
- assists synchronous learners in resurfacing their knowledge after the course is complete
- links important context with knowledge being shared
- makes use of a primary communication media

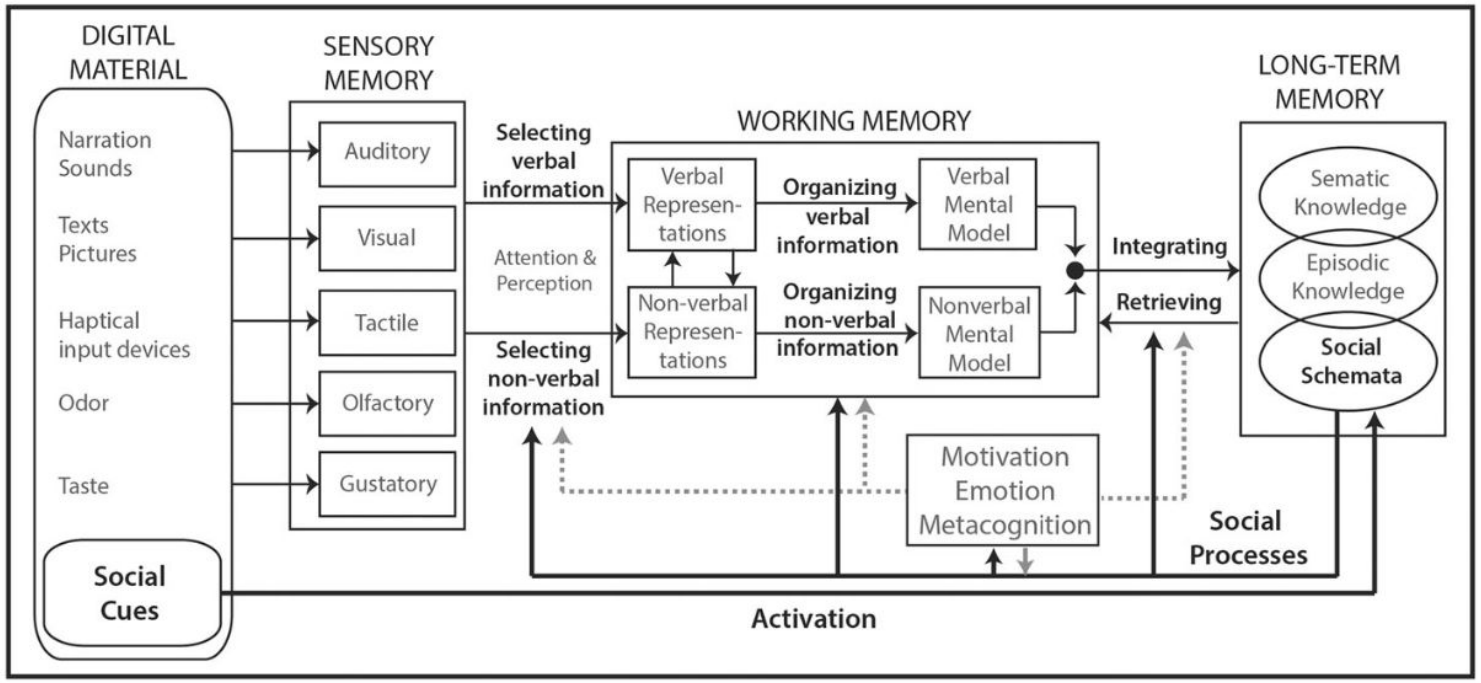


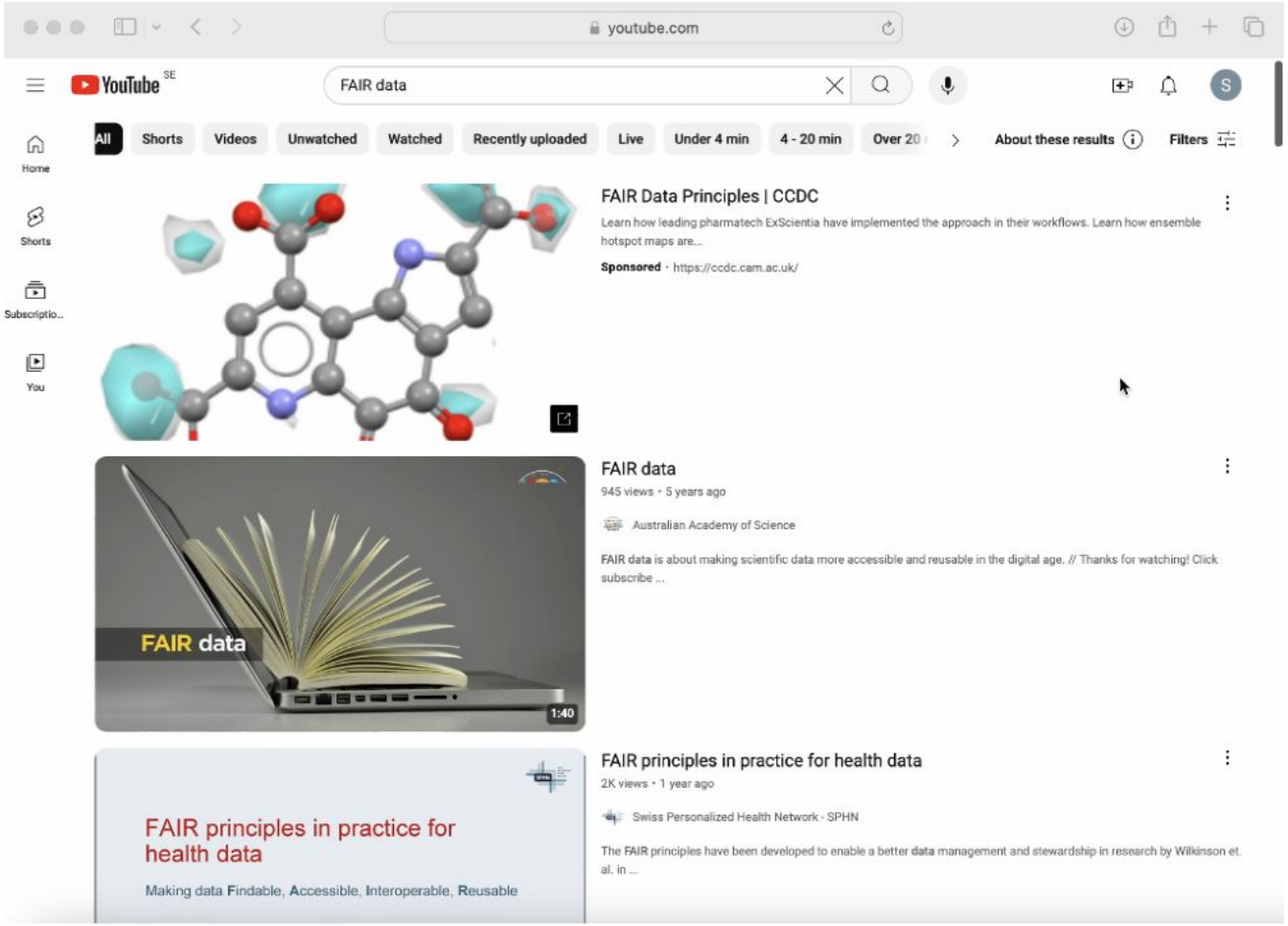
Fig. 1 The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE) with emphasis on the mediation of social processes on selection, organization, integration, and retrieval processes. The CASTLE is based on the Cognitive-Affective Theory of Learning with Media according to Moreno (2006)

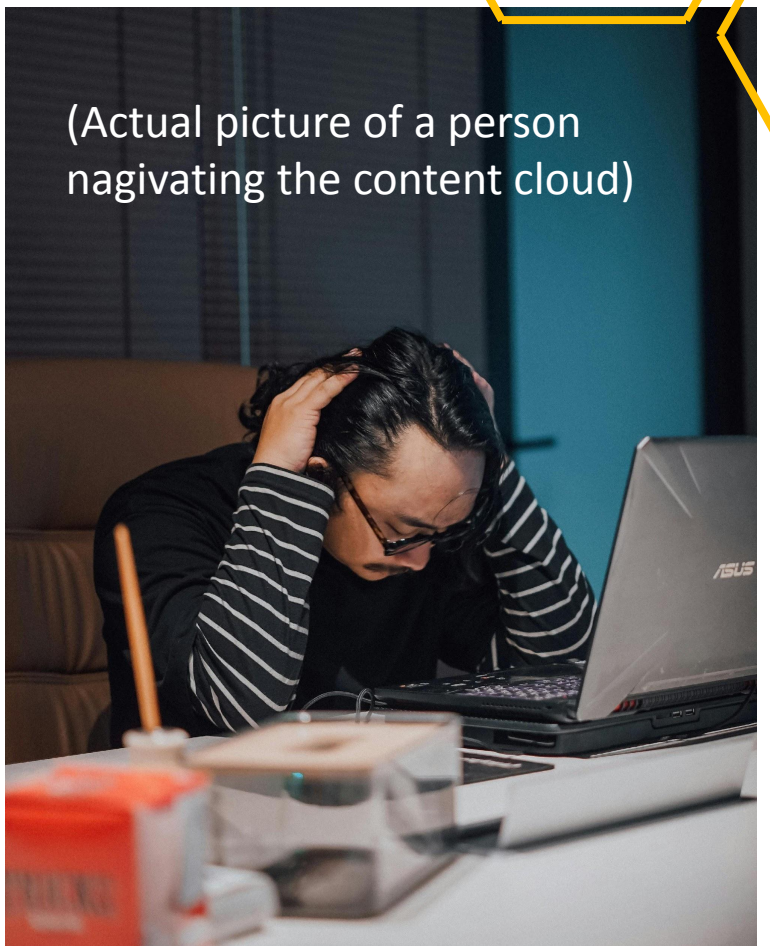
Video media is a great tool for **enhancing & contextualizing learning**!

Part 1: Why do we care about having FAIR videos?

video content ≠ learning material

however…





Your video learning material will exist within the content cloud

Part 1: Why do we care about having FAIR videos?

if video content ≠ learning material

what's the difference?

Zoom recording of my amazing lecture!!!!

Like and subscribe :D

IL **Cool science guy**

Constructive alignment with discussions and exercises

Metadata record

Where to find expert input or community for schema testing

Online course delivery

Explanation of how stress influences protein folding by a PhD who studied it in a lab for 5 years

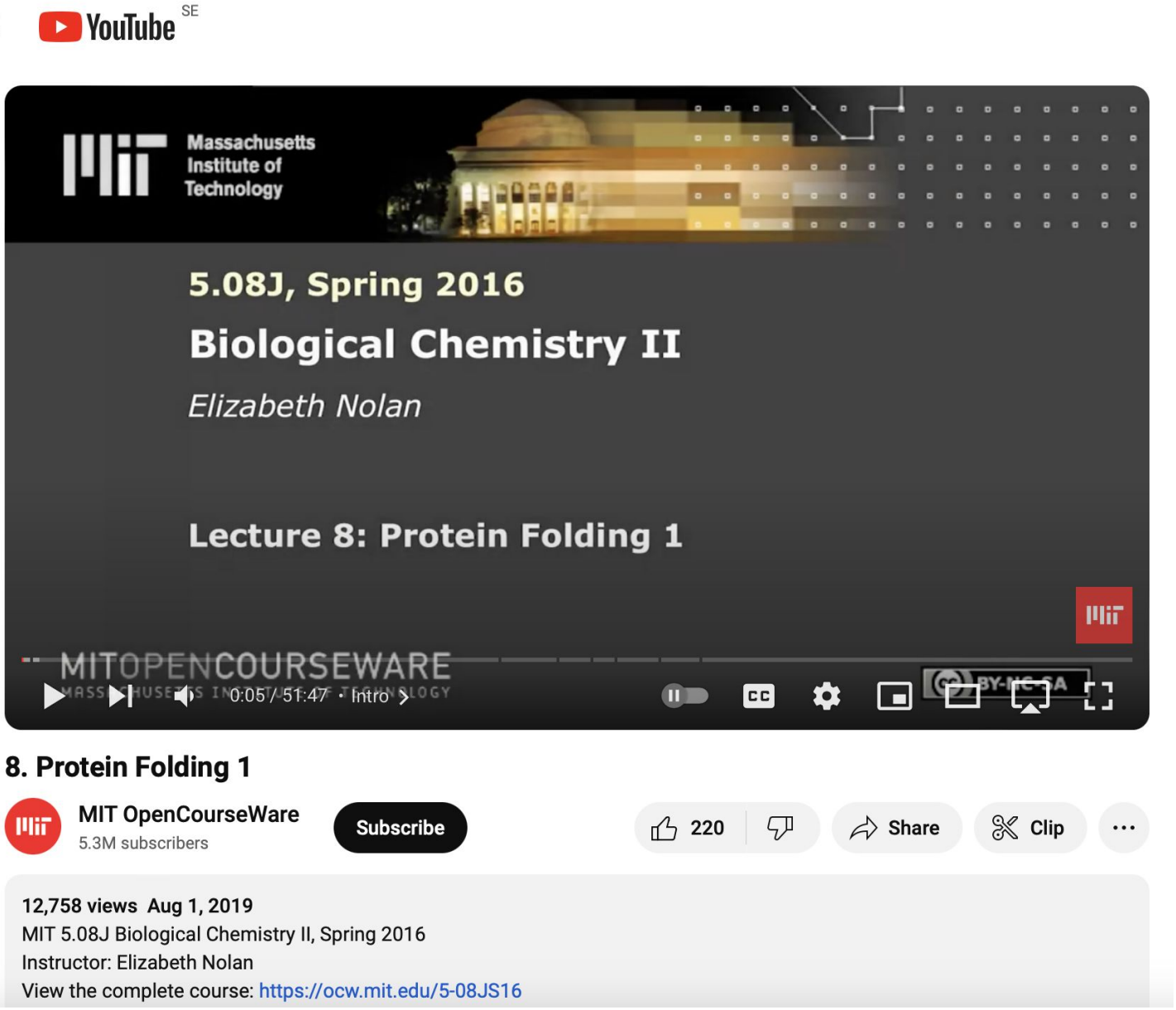
SciLifeLab Training Hub

Doi: xxx
Author: Cool science guy

Transcript:
This lecture is a part of the course Stress 101 delivered in December 2034 at Stockholm University. Hello and welcome to this short lecture on…

Video learning materials need **FAIR metadata** to place them in the appropriate **context** to be understood by **both humans and machines**.

Part 1: Example #1 – how FAIR is it?



Link

Part 1: Example #2 – how FAIR is it?

protocols.io SEARCH SIGN IN SIGN UP

Nov 15, 2020 0

Video protocol for sorting Drosophila pupae

In 1 collection

DOI
dx.doi.org/10.17504/protocols.io.bpq3mmyn

Carolyn Elya[1]
[1]Harvard University

Carolyn Elya
Harvard University

1 1 RUN COPY / FORK

Steps Metadata Metrics

Abstract

This is my favorite way to sort pupae.

Credit to Tom Alisch & Dave Zucker (@FlySorter) for discovering a new use for transparencies!!

1 Video summary of the method:

Link

Part 1: Minimum guidelines for integrating FAIR video materials

1. Findable

- Put the video link in the DOI record and vice versa
- Choose your video host thoughtfully, considering target audience & longevity

2. Accessible

- Ensure access type and language are defined in the hosting metadata
- Create a transcript or summary of the video for longevity, or consider hosting a download of your material

3. Interoperable

- Add a transcript to your DOI record - .srt, .txt
- Use common containers for downloadable content - .mp4, .mov, etc.

4. Reusable

- Choose a license for your material – remixing is a common term on video hosting platforms
- Understand requirements for tagging AI content

Part 2: Creating FAIR video learning materials

Minimum guidelines

Creating FAIR video learning materials

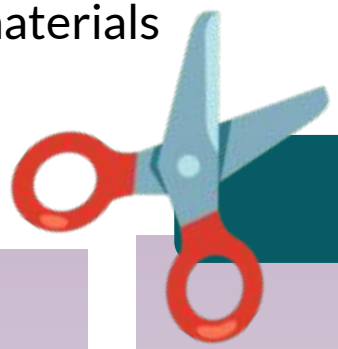
Video production is beyond the scope of this course :(

Part 2: Creating FAIR video learning materials

Create

Capture:
- Zoom
- Powerpoint
- Your smartphone
- OBS Studio

Open material:
- Pexels
- Pixabay



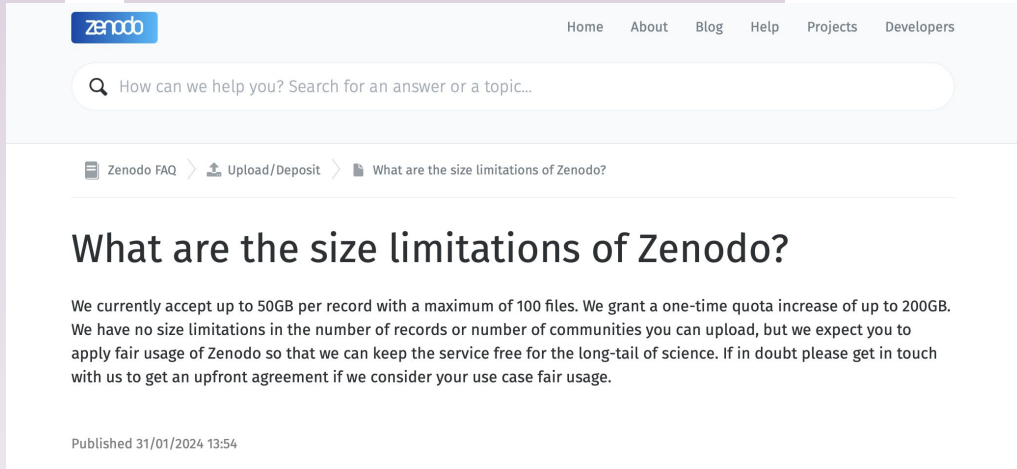
Edit

Open source:
- Blender
- Kdenlive
- OBS Studio

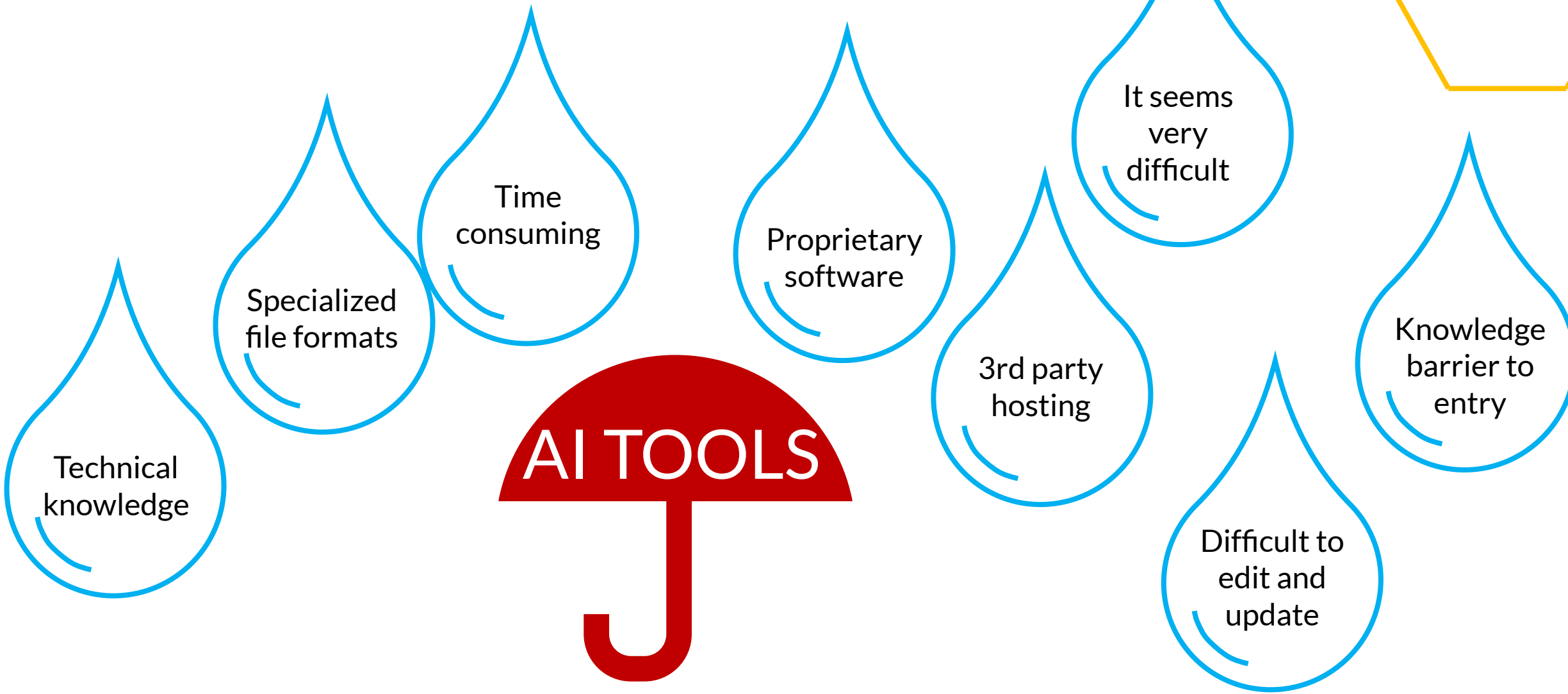
Proprietary:
- iMovie
- Adobe Premiere
- Many others

Release

Hosting options:
- YouTube
- Vimeo
- Nextcloud
- Google Drive
- Zenodo*



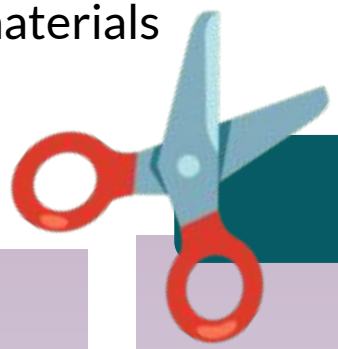
Barriers to creating FAIR video learning material include **time, skill, knowledge,** and **control**.



Create

Capture:
- **Zoom**
- **Powerpoint**
- **Your smartphone**
- OBS Studio

Open license:
- Pexels
- **Pixabay**



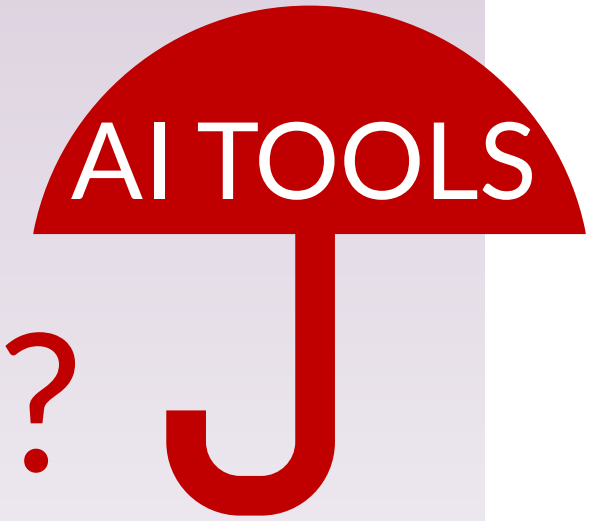
Edit

Open source:
- Blender
- Kdenlive
- OBS Studio

Proprietary:
- **iMovie**
- **Adobe Premiere**
- **Many others**

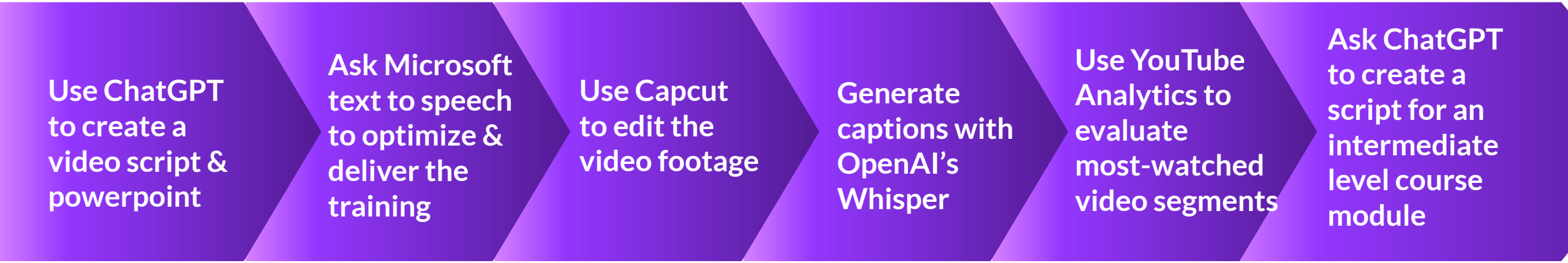
Release

Hosting options:
- **YouTube**
- **Vimeo**
- Nextcloud
- **Google Drive**
- Zenodo*

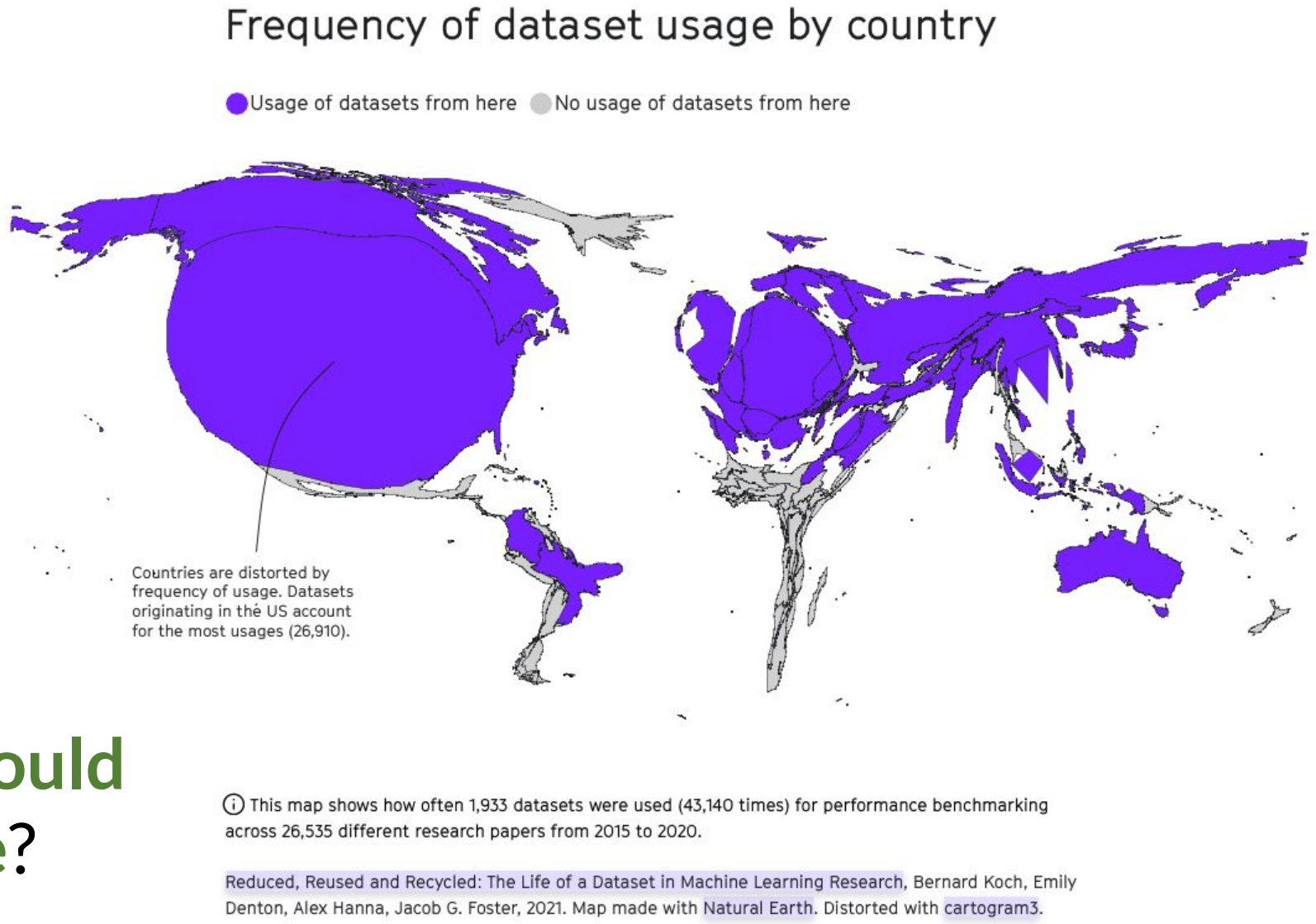


So what?

Part 2: Discussion – automating education



- Is the output of AI tools consistent with FAIR principles to improve discoverability and reproducibility?
- Need active and responsible choice in using AI tools within higher education & lifelong learning
- Normative vs. descriptive view

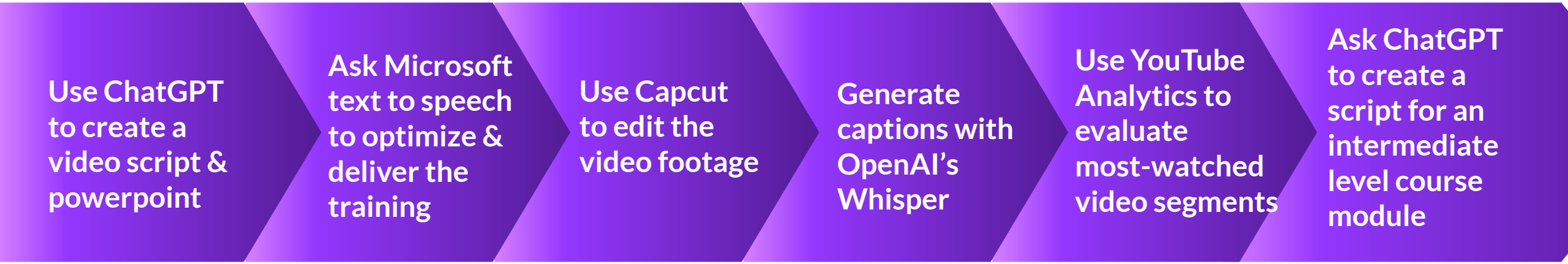


Guiding question: **Should AI tools be used here**?

Image: Mozilla internet health report

Slide adapted from Virginia Dignum lecture – Beyond the AI hype: balancing innovation and social responsibility

Part 3: Discussion – automating education



How might using AI tools like this affect learning?

What biases might this genAI output be subject to?

Where would you not want to use AI tools?

Where could AI tools help you in developing learning materials?

Part 4: Exercise – 20mins

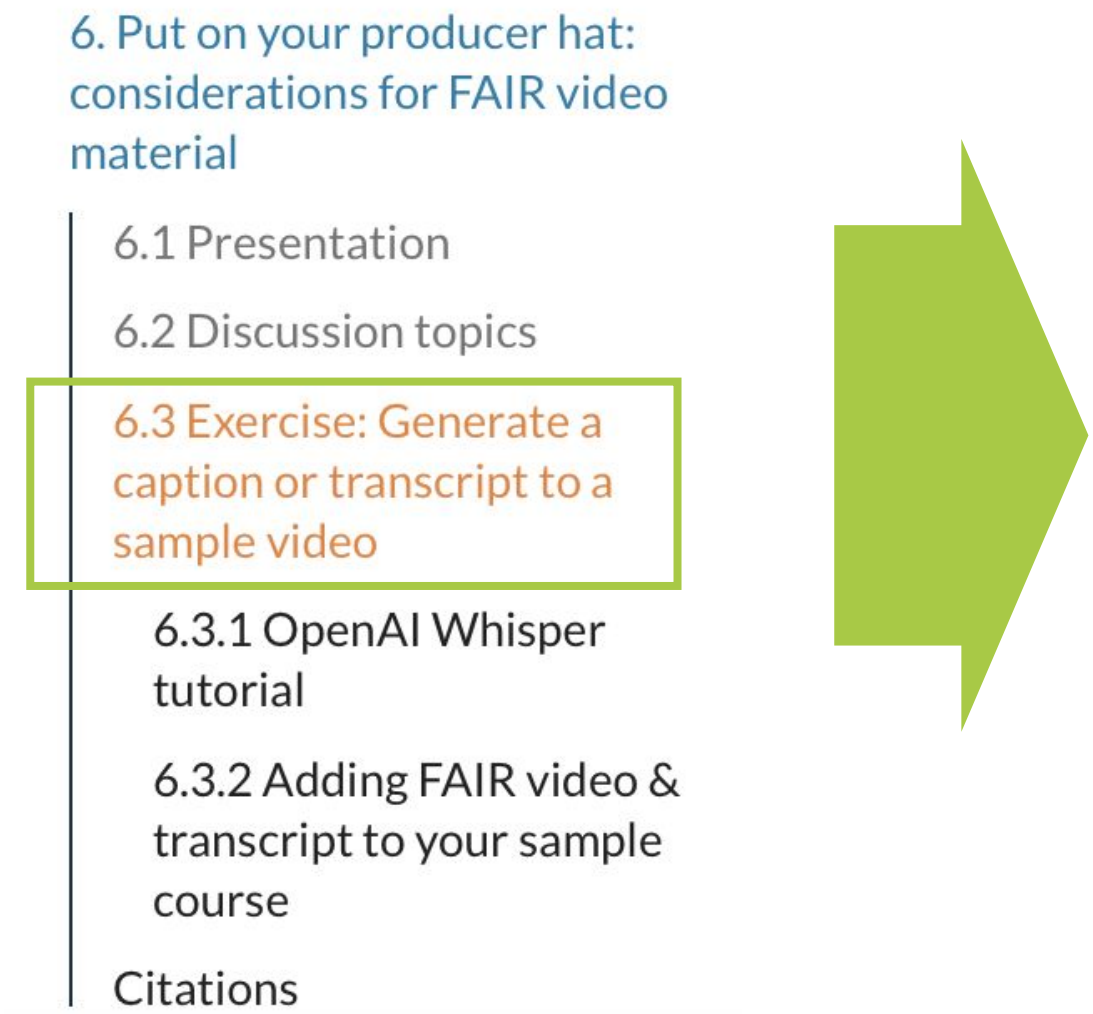
Lets add some **FAIR videos** to our training!

Refer to the course website (exercise 6.3) and choose your own adventure

Toast mode: Sample video

Konditori mode: Film yourself

Part 4: Exercise – improve FAIRness by generating a transcript



Recommended route:

During the course:

1. Navigate to the whisper container
   - https://whisper-ai.serve.scilifelab.se/
2. Upload your video file
3. Select the output format(s) - .txt for transcript and .srt for captions
4. Submit the file

When the file has completed transcription, the output will download to your computer.

Part 4: Exercise – improve FAIRness by generating a transcript

Check your transcript!
The word error rate (WER) for the **large** Whisper model is 4.1% for English and 7.6% for Swedish.

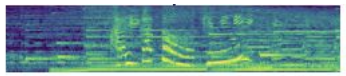
- Is it accurate?
- Is it usable?
- How does it help me achieve FAIR learning aims?

Part 4: Exercise – what happened?

What happens under the hood :: HTTPS://DOI.ORG/10.48550/ARXIV.2212.04356

audio input

30s chunks of audio are transformed to digital data



Log-Mel Spectrogram

Translates the spectrogram into numerical representation

encoder

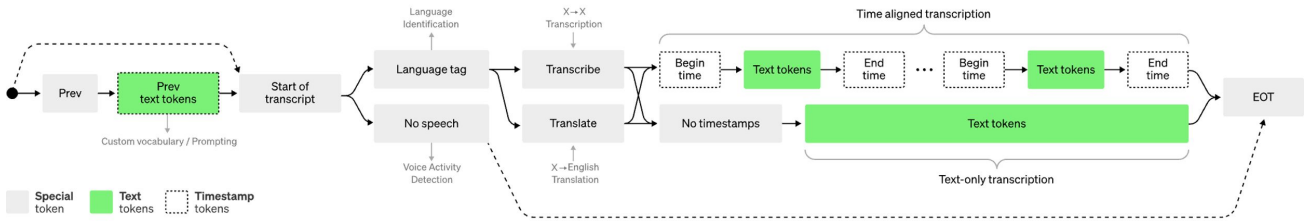
decoder

Predicts the words based on the training data sets* and produces the output

```
The quick
brown fox…
```

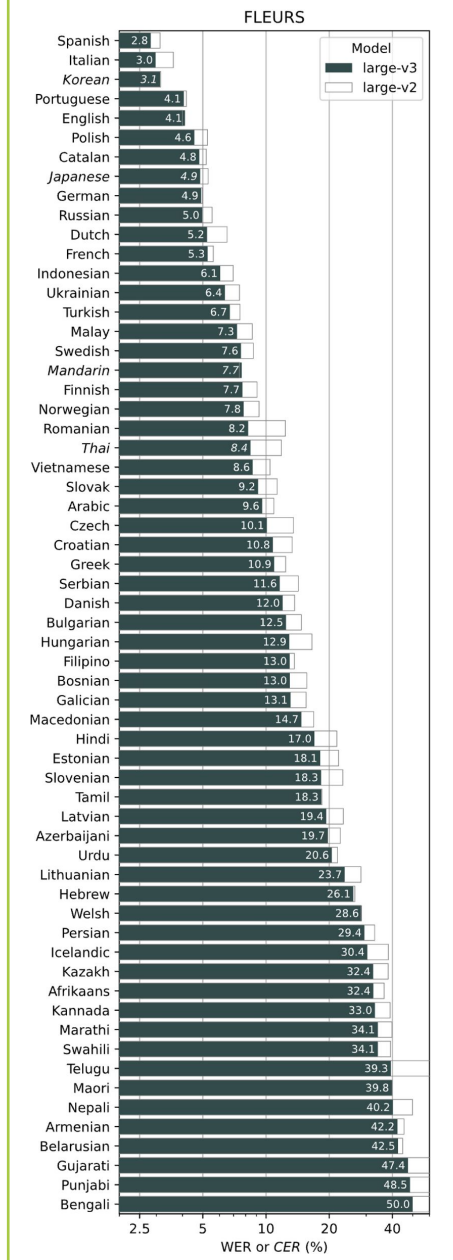
Data sets: scraped web text, YouTube video transcriptions, podcasts[1]

This process is repeated for the length of the audio:



Sources:

1. https://www.theverge.com/2024/4/6/24122915/openai-youtube-transcripts-gpt-4-training-data-google
2. https://web.archive.org/web/20230820005801/https://openai.com/research/whisper
3. https://github.com/openai/whisper



Part 4: Exercise

Integrate a video into your course page:

1. You can use the sample video on youtube to embed
    - Feel free to upload your own video to your host of choice, but we will not have time to watch these in the presentations ☺️
2. Attribute the video or choose a license
3. Add your caption file!

- Algorithms of Oppression - Safiya Umoja Noble
    - Extended reading: An Intersectional Physics Identity Framework for Studying Physics Settings – Angela Johnson
- Artificial Intelligence and Learning Futures - Stefan Popenici
- Beyond the AI hype: balancing innovation and social responsibility - Virginia Dignum